I.Input, output của bài toán:

1.Input:

* Gồm 16 file csv chứa dữ liệu về cổ phiếu của 16 ngân hàng ở Việt Nam.
* 16 ngân hàng là:

1.CTCP Bảo hiểm Ngân hàng Nông nghiệp Việt Nam - ABI - 3083

2.Ngân hàng TMCP Á Châu - ACB - 3740

3.CTCP Chứng khoán Agribank - AGR - 2978

4.Tổng Công ty cổ phần Bảo Minh - BMI - 3726

5.Tổng Công ty cổ phần Bảo hiểm Ngân hàng Đầu tư và Phát triển Việt Nam - BIC - 2548

6.Ngân hàng TMCP Đầu tư và Phát triển Việt Nam - BID - 1948

7.Tập đoàn Bảo Việt - BVH - 3098

8.Ngân hàng TMCP Công Thương Việt Nam - CTG - 3083

9.Ngân hàng TMCP Xuất nhập khẩu Việt Nam - EIB - 3010

10.Ngân hàng TMCP Quân Đội - MBB - 2508

11.Ngân hàng TMCP Quốc Dân - NVB - 2794

12.Ngân hàng TMCP Ngoại thương Việt Nam - VCB - 3510

13.Ngân hàng TMCP Sài Gòn - Hà Nội - SHB - 3832

14.Ngân hàng TMCP Sài Gòn Thương Tín - STB - 3096

15.Tổng Công ty cổ phần Bảo hiểm Bưu điện - PTI - 2679

16.CTCP PVI - PVI – 3563

* Mỗi file csv chứa thông tin:

1.Ticker: tên cổ phiếu

2. DTYYYYMMDD: ngày / tháng /năm

3.OpenFixed: Giá mở bán trong ngày

4.HighFixed: Giá cao nhất trong ngày

5.LowFixed: Giá thấp nhất trong ngày

6.CloseFixed: Giá đóng cửa

7.Volume: Khối lượng giao dịch

2.Output

* Output của bài toán sẽ gồm các thông tin sau:

Previous\_total\_asset:Tổng tài sản ban đầu

End\_total\_asset:Tổng tài sản cuối cùng

Total\_reward: tổng phần thưởng

Total\_cost: tổng chi phí

Total trades: tổng số giao dịch

Sharpe: đo xem lợi nhuận thu được là bao nhiêu trên một đơn vị rủi ro khi đầu tư vào một tài sản hay đầu tư theo một chiến lược kinh doanh.

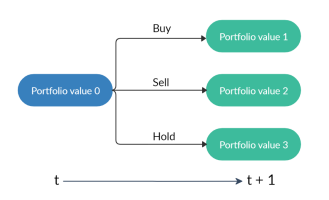
II.PROBLEM DESCRIPTION

Mô hình hóa giao dịch chứng khoán bằng cách sử dụng Quy trình Quyết định Markov (MDP), và xây dựng mục tiêu giao dịch là tối đa hóa lợi nhuận kỳ vọng.

1. MDP model cho giao dịch chứng khoán

Để mô hình hóa bản chất ngẫu nhiên của chứng khoán động thị trường, chúng tôi sử dụng MDP như sau:

* State s = [p, h, b]: một vectơ bao gồm giá cổ phiếu p ∈ , cổ phiếu chứng khoán h ∈ và số dư còn lại b ∈ R+ , trong đó D là số lượng cổ phiếu, Z+­ là số nguyên không âm
* Action a: 1 vecto hành động trên D cổ phiếu. Các hành động cho phép trên mỗi cổ phiếu có quyền mua, bán ,giữ mà kết quả tăng, giảm, không đổi của cổ phiếu h.
* Reward r (s, a, s0): phần thưởng trực tiếp của action a tại state s và chuyển đến new\_state s
* Policy π (s): chiến lược giao dịch tại state s, là phân phối xác suất của các hành động ở state s.
* Q-value Q π (s, a): phần thưởng mong đợi khi thực hiện action a ở state s tuân theo policy π.
* +, và phần còn lại
* số dư b ∈ R +, trong đó D là số
* cổ phiếu và Z + biểu thị các số nguyên không âm.



Quá trình chuyển đổi trạng thái của một quá trình giao dịch chứng khoán

Tại mỗi trạng thái, một trong 3 action có thể thực hiện là lấy trên cổ phiếu d (d = 1, ..., D) trong danh mục đầu tư.

* Sell: k [d] ∈ [1, h [d]] cổ phiếu dẫn đến ht+1[d] = ht[d] - k[d], trong đó

k[d] ∈ Z + và d = 1, ..., D.

* Hold: ht+1[d] = ht[d]
* Buy: k[d] cổ phiếu dẫn đến ht+1[d] = ht[d] + k[d],

Tại thời điểm t có hành động nào được thực hiện và giá cổ phiếu cập nhật tại t+1.Do đó portfolio values có thể thay đổi từ giá trị này sang giá trị khác. Portfolio values là pTh+b

1. Kết hợp các ràng buộc giao dịch chứng khoán

Giả định và ràng buộc sau đây phản ánh các điều kiện thực tế: chi phí giao dịch, thanh khoản thị trường, ngại rủi ro, v.v.

* Tính thanh khoản của thị trường: các lệnh có thể được thực hiện nhanh chóng tại giá đóng cửa. Chúng tôi giả định rằng thị trường chứng khoán sẽ không bị ảnh hưởng bởi đại lý kinh doanh cốt thép của chúng tôi
* Chi phí giao dịch: chi phí giao dịch được phát sinh cho mỗi giao dịch. Có nhiều loại chi phí giao dịch chẳng hạn như phí trao đổi, phí thực hiện và phí SEC. Các nhà môi giới khác nhau có phí hoa hồng khác nhau. Mặc dù có những khác biệt về phí, chúng tôi giả định chi phí giao dịch của chúng tôi là 0,1% giá trị của mỗi giao dịch
* Không thích rủi ro đối với sự sụp đổ của thị trường: có những các sự kiện có thể gây ra sự sụp đổ của thị trường chứng khoán, chẳng hạn như chiến tranh, sự sụp đổ của bong bóng thị trường chứng khoán, nợ có chủ quyền

vỡ nợ và khủng hoảng tài chính.

1. Tối đa hóa lợi nhuận làm mục tiêu giao dịch

* Xác định reward function là sự thay đổi của portfolio value khi action a được thực hiện ở state s và đến new\_state s. Mục tiêu là thiết kế một chiến lược giao dịch tối đa hóa sự thay đổi của portfolio value:

r (st, at, st+1) = (bt+1+ t+1) – (bt + ht) – ct

* Mô hình được khởi tạo như sau. p0 là giá cổ phiếu tại thời điểm 0 và b0 là số vốn ban đầu. h và Q π (s, a) là 0, và π (s) được phân bố đồng đều giữa tất cả các hành động cho mỗi trạng thái. Sau đó, Qπ (st, at) được cập nhật thông qua tương tác với môi trường thị trường chứng khoán. Chiến lược tối ưu được đưa ra bởi Phương trình Bellman rằng phần thưởng mong đợi khi thực hiện action at ở state st là kỳ vọng tổng kết của phần thưởng trực r (st, at, st+1) và phần thưởng trong tương lai ở trạng thái tiếp theo st+1. Phần thưởng trong tương lai được chiết khấu theo hệ số 0 <γ <1 cho mục đích hội tụ.

Q π (st, at) = Es(t+1)[ r (st, at, st+1) + γ Ea(t+1) ~ π(st+1)[Qπ(st+1,at+1)]].

* Mục tiêu là thiết kế một chiến lược giao dịch tối đa hóa sự thay đổi tích lũy tích cực của danh mục đầu tư giá trị r (st, at, st+1) trong môi trường năng động và chúng tôi sử dụng phương pháp the deep reinforcement learning để giải quyết vấn đề này.